

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-112967

(43)Date of publication of application : 21.04.2000

(51)Int.Cl.

G06F 17/30

G06F 17/28

(21)Application number : 10-280733

(71)Applicant : NIPPON TELEGR & TELEPH CORP
<NTT>

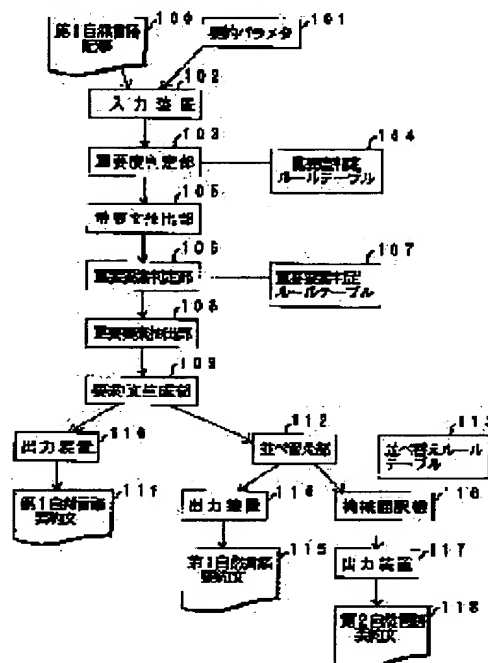
(22)Date of filing : 02.10.1998

(72)Inventor : HATAYAMA MAMIKO
SHIRAI SATOSHI(54) ARTICLE SUMMARIZATION DEVICE, ARTICLE SUMMARIZATION PROCESSING METHOD
AND RECORD MEDIUM RECORDING PROGRAM DESCRIBING THE METHOD

(57)Abstract:

PROBLEM TO BE SOLVED: To automatically summarize an article in a desirable form by providing an important element extraction means, a summary sentence generation means which generates a summary sentence based on the extracted important elements, an output means which outputs the generated summary sentence, etc.

SOLUTION: The importance is given to every element via an important element decision part 106 and by means of an important element decision rule table 107. Then the important elements are extracted via an important element extraction part 108 against the result that is obtained by giving important to every element. In other words, the part 108 extracts the subject, the verb and other modifiers of each of several sentences to which the importance is given to every element. These extracted important elements and the elements of the extracted subject, verb, etc., are turned into a summary sentence 111 via a summary sentence generation part 109. Then the sentence 111 is outputted by an output device 110. Thus, the sentence 111 is obtained.



LEGAL STATUS

[Date of request for examination] 21.12.2000

[Date of sending the examiner's decision of rejection] 15.06.2004

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

BEST AVAILABLE COPY

[Number of appeal against examiner's decision
of rejection]

[Date of requesting appeal against examiner's
decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2000-112967
(P2000-112967A)

(43) 公開日 平成12年4月21日 (2000. 4. 21)

(51) Int.Cl. ⁷	識別記号	F I	テーマコード* (参考)
G 0 6 F 17/30		G 0 6 F 15/401	3 2 0 A 5 B 0 7 5
17/28		15/38	Q 5 B 0 9 1

審査請求 未請求 請求項の数9 O L (全 8 頁)

(21) 出願番号 特願平10-280733

(22) 出願日 平成10年10月2日 (1998. 10. 2)

(71) 出願人 000004226

日本電信電話株式会社
東京都千代田区大手町二丁目3番1号

(72) 発明者 畑山 満美子

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(72) 発明者 白井 諭

東京都新宿区西新宿三丁目19番2号 日本
電信電話株式会社内

(74) 代理人 100087848

弁理士 小笠原 吉義 (外1名)

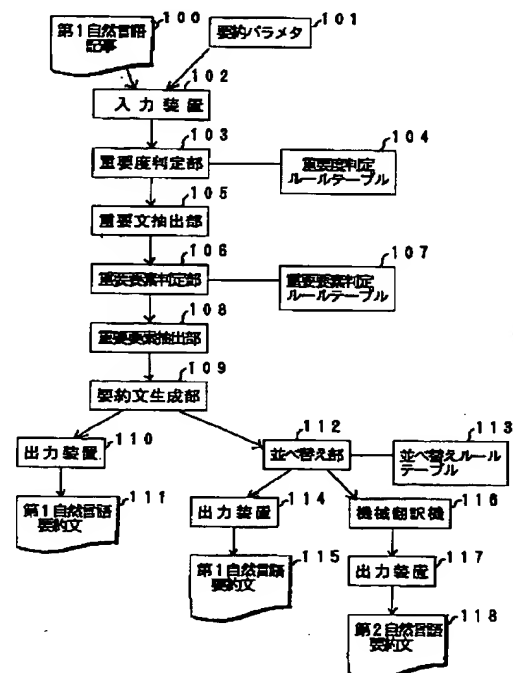
Fターム(参考) 5B075 ND03 NS01 PQ20 UU06
5B091 AA05 AA15 BA03 CA02 CA05
CD03 CD13 EA00

(54) 【発明の名称】 記事要約装置および記事要約処理方法および当該方法を記述したプログラムを記録した記録媒体

(57) 【要約】

【課題】 本発明は、重要度判定ルールテーブルや重要要素判定ルールテーブルを用いて好ましい形で記事要約を自動的に行うようにすることを目的としている。

【解決手段】 重要度判定ルールテーブルの内容を用いて一文ごとに重要度を判定する手段と、重要要素判定ルールテーブルの内容を用いて重要要素を判定する手段とを用いて、与えられた記事を要約する。



【特許請求の範囲】

【請求項 1】 記事要約を行なう記事要約装置において、
記事を読み込み、かつ、記事全体および要素の双方に対する、少なくとも、記事短縮度合いと重要度閾値とを指定する要約パラメータを読み込む入力手段と、
重要度判定ルールテーブルと、
入力された記事に対し、上記重要度判定ルールを用いて、一文ごとに重要度を判定する重要度判定手段と、
上記の重要度判定結果と前記記事全体に対する要約パラメータとに基づいて重要文を抽出する重要文抽出手段と、
重要要素判定ルールテーブルと、
抽出された重要文に対し、上記重要要素判定ルールを用いて、重要要素を判定する重要要素判定手段と、
上記の重要要素判定結果と前記要素に対する要約パラメータとに基づいて重要要素を抽出する重要要素抽出手段と、
上記の抽出された重要要素に基づいて要約文を生成する要約文生成手段と、
生成された要約文を出力する出力手段とを有することを特徴とする記事要約装置。

【請求項 2】 記事要約装置において、
並べ替えルールテーブルと、
要約文生成手段の出力に対し、上記並べ替えルールを用いて並べ替えを行う並べ替え手段と、
並べ替えられた要約文を出力する出力手段とを有することを特徴とする請求項（1）記載の記事要約装置。

【請求項 3】 記事要約装置において、
第 1 自然言語で出力された並べ替え手段の出力に対し、
第 2 自然言語に機械翻訳を行う機械翻訳手段と、
第 2 自然言語に翻訳された要約文を出力する出力手段とを有することを特徴とした請求項（2）記載の記事要約装置。

【請求項 4】 記事要約を行なう記事要約処理方法において、
記事を読み込み、かつ、記事全体および要素の双方に対する、少なくとも、記事短縮度合いと重要度閾値とを指定する要約パラメータを読み込み、
重要度判定ルールテーブルを用意しておいて、入力された記事に対し、上記重要度判定ルールを用いて、一文ごとに重要度を判定し、
上記の重要度判定結果と前記記事全体に対する要約パラメータとに基づいて重要文を抽出し、
重要要素判定ルールテーブルを用意しておいて、抽出された重要文に対し、上記重要要素判定ルールを用いて、重要要素を判定し、
上記の重要要素判定結果と前記要素に対する要約パラメータとに基づいて重要要素を抽出し、
上記の抽出された重要要素に基づいて要約文を生成し、
生成された要約文を出力することを特徴とする記事要約

処理方法。

【請求項 5】 記事要約処理方法において、
並べ替えルールテーブルを用意しておいて、要約文生成手段の出力に対し、上記並べ替えルールを用いて並べ替えを行い、
並べ替えられた要約文を出力することを特徴とする請求項（4）記載の記事要約処理方法。

【請求項 6】 記事要約処理方法において、
第 1 自然言語で出力された並べ替え手段の出力に対し、
第 2 自然言語に機械翻訳を行い、
第 2 自然言語に翻訳された要約文を出力することを特徴とした請求項（5）記載の記事要約処理方法。

【請求項 7】 記事要約を行なう記事要約処理方法を記述したプログラムを記録した記録媒体において、
記事を読み込み、かつ、記事全体および要素の双方に対する、少なくとも、記事短縮度合いと重要度閾値とを指定する要約パラメータを読み込み、
重要度判定ルールテーブルを用意しておいて、入力された記事に対し、上記重要度判定ルールを用いて、一文ごとに重要度を判定し、
上記の重要度判定結果と前記記事全体に対する要約パラメータとに基づいて重要文を抽出し、
重要要素判定ルールテーブルを用意しておいて、抽出された重要文に対し、上記重要要素判定ルールを用いて、重要要素を判定し、
上記の重要要素判定結果と前記要素に対する要約パラメータとに基づいて重要要素を抽出し、
上記の抽出された重要要素に基づいて要約文を生成し、
生成された要約文を出力する記事要約処理方法について、
当該記事要約処理方法を記述したプログラムを記録したことを特徴とする記録媒体。

【請求項 8】 記事要約処理方法を記述したプログラムを記録した記録媒体において、
請求項（7）記載の記事要約処理方法に付加して、
並べ替えルールテーブルを用意しておいて、要約文生成手段の出力に対し、上記並べ替えルールを用いて並べ替えを行い、
並べ替えられた要約文を出力する記事要約処理方法について、
当該記事要約処理方法を記述したプログラムを記録したことを特徴とする請求項（7）記載の記録媒体。

【請求項 9】 記事要約処理方法を記述したプログラムを記録した記録媒体において、
請求項（8）記載の記事要約処理方法に付加して、
第 1 自然言語で出力された並べ替え手段の出力に対し、
第 2 自然言語に機械翻訳を行い、
第 2 自然言語に翻訳された要約文を出力する記事要約処理方法について、
当該記事要約処理方法を記述したプログラムを記録した

ことを特徴とする請求項（８）記載の記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】本発明は、第１自然言語で記述された記事の要約や第２自然言語に翻訳された要約を行なう装置および記事要約処理方法および当該方法を記述したプログラムを記録した記録媒体に関する。

【0002】

【従来の技術】従来の要約技術では、人が指定したキーワードの重み付けに従って、要約装置が重みによる点数付けによって重要文を判定し、重要文を抽出する手法、単語の出現回数を統計的に計算し、重要文を抽出する方法、重要文選択に関する心理実験を利用する手法（複数の被験者に同一文章を要約させ、その結果を統計処理しキーワードの自動重み付けを行ない、それにより要約装置に反映させ、要約を行なう方法がある）、などが知られている。

【0003】これらに関しては、奥村学「テキスト自動要約技術の現状と課題」(<http://www.jaist.ac.jp/oku/okumura-j.html>)に詳しい。

【0004】

【発明が解決しようとする課題】従来の要約技術では、人が指定したキーワードの重み付けに従って、要約装置が重みによる点数付けによって重要文を判定し、重要文を抽出する手法では、人手でキーワードの重み付けを行なうため、人的時間的コストがかかる問題がある。

【0005】また、単語の出現回数を統計的に計算し、重要文を抽出する方法では、抽出した文をそのまま並べた場合、意味がつかないという問題がある。

【0006】また、複数の被験者に同一文章を要約させ、その結果を統計処理し、キーワードの自動重み付けを行ない、それにより要約装置に反映させ、要約を行なう方法では、要約結果の評価が曖昧であることと、評価データを作成するコストがかかる問題がある。

【0007】また、従来の技術では、記事全体の５割程を重要文として取り出す程度の要約しか行なえない。

【0008】また、従来の技術では、他言語に要約する技術は考えられていなかった。といった問題があった。

【0009】本発明は、これらの問題を解決するものであり、重要度判定ルールテーブルや重要要素判定ルールテーブルを用いて好ましい形で記事要約を自動的に行うようにすることを目的としている。

【0010】

【課題を解決するための手段】本発明では、請求項

（１）の記事と、記事全体および要素の双方に対する記事短縮度合い、重要度閾値を指定する要約パラメタを読み込む入力手段と、重要度判定ルールテーブルと、入力された記事に対し、上記重要度判定ルールを用いて、一文ごとに重要度を判定する重要度判定手段と、上記の重要度判定結果と前記記事全体に対する要約パラメタとに

基づいて重要文を抽出する重要文抽出手段と、重要要素判定ルールテーブルと、抽出された重要文に対し、上記重要要素判定ルールを用いて、重要要素を判定する重要要素判定手段と、上記の重要要素判定結果と前記要素に対する要約パラメタとに基づいて重要要素を抽出する重要要素抽出手段と、上記の抽出された重要要素に基づいて要約文を生成する要約文生成手段と、生成された要約文を出力する出力手段とを有することを特徴とする請求項（１）記載の記事要約装置によって、キーワードによる重み付けや統計処理を行なうコストを省き、文レベルではなく、要素レベルでの要約を可能とし、予め与えた要約パラメタにより要約度を調整できることを可能とする。

【0011】また、本発明では、請求項（２）の記事と、記事全体および要素の双方に対する記事短縮度合い、重要度閾値を指定する要約パラメタを読み込む入力手段と、重要度判定ルールテーブルと、入力された記事に対し、上記重要度判定ルールを用いて、一文ごとに重要度を判定する重要度判定手段と、上記の重要度判定結果と前記記事全体に対する要約パラメタとに基づいて重要文を抽出する重要文抽出手段と、重要要素判定ルールテーブルと、抽出された重要文に対し、上記重要要素判定ルールを用いて、重要要素を判定する重要要素判定手段と、上記の重要要素判定結果と前記要素に対する要約パラメタとに基づいて重要要素を抽出する重要要素抽出手段と、上記の抽出された重要要素に基づいて要約文を生成する要約文生成手段と、生成された要約文を出力する出力手段と、並べ替えルールテーブルと、要約文生成手段の出力に対し、上記並べ替えルールを用いて並べ替えを行う並べ替え手段と、並べ替えられた要約文を出力する出力手段とを有することを特徴とする請求項（２）記載の記事要約装置によって、キーワードによる重み付けや統計処理を行なうコストを省き、文レベルではなく、要素レベルでの要約を可能とし、予め与えた要約パラメタにより要約度を調整できることを可能とし、また、文および要素の並べ替えを行なうことによって、単なる文の抽出だけでは意味がつかないという問題を解決する。

【0012】また、本発明では、請求項（３）の記事と、記事全体および要素の双方に対する記事短縮度合い、重要度閾値を指定する要約パラメタを読み込む入力手段と、重要度判定ルールテーブルと、入力された記事に対し、上記重要度判定ルールを用いて、一文ごとに重要度を判定する重要度判定手段と、上記の重要度判定結果と前記記事全体に対する要約パラメタとに基づいて重要文を抽出する重要文抽出手段と、重要要素判定ルールテーブルと、抽出された重要文に対し、上記重要要素判定ルールを用いて、重要要素を判定する重要要素判定手段と、上記の重要要素判定結果と前記要素に対する要約パラメタとに基づいて重要要素を抽出する重要要素抽出

手段と、上記の抽出された重要要素に基づいて要約文を生成する要約文生成手段と、生成された要約文を出力する出力手段と、並べ替えルールテーブルと、要約文生成手段の出力に対し、上記並べ替えルールを用いて並べ替えを行う並べ替え手段と、並べ替えられた要約文を出力する出力手段と、第1自然言語で出力された並べ替え手段の出力に対し、第2自然言語に機械翻訳を行う機械翻訳手段と、第2自然言語に翻訳された要約文を出力する出力手段とを有することを特徴とする請求項(3)記載の記事要約装置によって、キーワードによる重み付けや統計処理を行なうコストを省き、文レベルではなく、要素レベルでの要約を可能とし、予め与えた要約パラメタにより要約度を調整できることを可能とし、文および要素の並べ替えを行なうことによって、単なる文の抽出だけでは意味がつかないという問題を解決し、従来の技術では、他言語に要約する技術が与えられていなかったという問題を解決する。

【0013】本発明では、キーワードによる重み付けや統計処理を行なうコストを省き、文レベルではなく、要素レベルでの要約を可能とし、予め与えた要約パラメタにより要約度を調整できることを可能とし、文および要素の並べ替えを行なうことによって、意味の通る要約文を生成することを可能とし、他言語に要約することを可能とする。

【0014】

【発明の実施の形態】第1自然言語で記述された記事の要約や第2自然言語に翻訳された要約を行なう装置およびその処理方法について、図2に示すある日本語新聞記事を要約する場合と、日本語新聞記事から英文に要約を行なう場合とを例にとって説明する。

【0015】図1は、本発明の記事要約装置の概略構成である。モジュール100～111は請求項(1)記載の記事要約装置に該当する構成図、モジュール100～109、112～115は請求項(2)記載の記事要約装置に該当する構成図、モジュール100～109、112～113、116～118は請求項(3)記載の記事要約装置に該当する構成図である。実施例を図2の日本語新聞記事を要約する場合を例として示す。

【0016】図1において、第1自然言語記事(日本語新聞記事)100と、記事全体および要素の双方に対する記事短縮度合い、重要度閾値を指定する要約パラメタ101が、入力装置102によって入力される。次に入力された新聞記事100に対し、図4で示す重要度判定ルールテーブル104の重要度判定ルールを用いて、図3で示す重要度判定部103によって、一文ごとに重要度を判定する。

【0017】なお、図3は重要度判定部における一実施例処理態様を示し、図4は重要度判定ルールテーブルの例を示す。

【0018】図3においては、手段301において記事

を1文ごとに分離し、手段302において重要度判定ルールテーブル303(図4参照)を利用して1文ごとに重要度を付与する。そして、手段304において1文ごとに重要度の付与された記事が出力される。このようにして、1文ごとに重要度の付与された記事が得られる。

【0019】また図4においては、例えば段落における文の順番や、固有名詞や、「～そうだ」などの表現を考慮して、重要度を定めている。

【0020】例えば、図2の新聞記事の場合、新聞記事は、重要度判定部103(図3)において、分割手段301によって、第1文から第5文の5文に分離された後、重要度付与手段302と重要度判定ルールテーブル303によって重要度が付与される。この例では、ルール401によって、第1文目>第2文目>第3文目>第4文目>第5文目の順に重要度が付与される。また、第1文は「日本たばこ産業」、第3文は「日本テレコム」、第4文は「JT株」、第5文は「NTT」「JR東日本」という企業名が出現するため、ルール402によって、第1、3、4、5文の重要度が高くなる。しかし、ルール404によって、出現順、つまり、第1文目>第3文目>第4文目>第5文目の順に重要度が付与される。これらの重要度付与により、第1文目>第2文目>第3文目>第4文目>第5文目の順に重要度が判定され、各文に重要度が付与された記事が出力手段304によって出力される。

【0021】次に、図1の重要文抽出部105によって、上記出力による各文に重要度が付与された記事は、予め与えられた要約パラメタ101のうちの記事全体に対する要約パラメタに従い優先度の高いものから任意数分の文が抽出される。重要文抽出部105の実施例を図5によって詳しく示す。図5において、手段501において、1文ごとに重要度が付与された記事と要約パラメタ502とから、記事全体に対する要約、パラメタに従って優先度の高いものから任意数分の文を抽出する。そして、手段503において、抽出された任意数分の文を出力する。

【0022】要約パラメタ502は、記事全体および要素の双方に対する記事短縮度合い、重要度閾値の4種類が入力される。ここでは、記事全体に対して6割程度、要素に対して7割程度に記事を短縮し、双方に対する重要度閾値を適度な値に設定するパラメタを入力した場合について説明する。記事全体に対して6割程度の圧縮率なので、図2に示す例文では、重要文抽出部501によって、上位3文(第1文、第2文、第3文)が抽出され、出力部503によって出力される。

【0023】次に、上記出力による3文は、図1の重要要素判定ルールテーブル107を用いて、重要要素判定部106によって、各要素に重要度が付与される。重要要素判定部106及び重要要素判定ルールテーブル107の実施例を、図6及び図7を用いて詳しく示す。図6

において、手段（解析部）601において先に抽出された任意数分の文について1文ごとに形態素解析や構文解析を行う。ついで手段（重要度付与部）602において重要要素判定ルールテーブル603（図7参照）を利用して1要素ごとに重要度を付与する。そして手段604において、1要素ごとに重要度が付与された任意数分の文を出力する。

【0024】また図7においては、名詞句や動詞について夫々重要度が定められている。

【0025】重要要素判定部106は、3文に抽出された前記出力に対し、解析部601によって、形態素解析および構文解析を行ない、文を要素単位に分割し、解析情報を得る。この解析情報と重要要素判定ルールテーブル603とを用いて、重要度付与部602によって、要素ごとに重要度を判定する。形態素解析および構文解析は、既存の技術を用いる。既存の技術については、例えば、長尾真編「自然言語処理」（岩波書店）に詳しい。

【0026】図2に示す例文の1文目では、「日本／たばこ／産業／（JT）：複合語（一般名詞）」「株／の：一般名詞／格助詞」「一般／売り出し／の：複合語（転生名詞）／格助詞」「購入／申し込み／が／、：用言性名詞／転生名詞／格助詞／文末記号」「八日／で：数詞／格助詞」「締め切り／れ／た／。：他動詞／助動詞／文末記号」というように形態素解析される。これと構文解析の結果、例文の1文目の主語は「購入申し込みが」、動詞は「締め切られた。」、また、目的語がないという情報が得られる。

【0027】これらの情報と、重要要素判定ルールテーブル603（図7に詳細）に基づいて重要度を判定すると、ルール701により、企業名「日本たばこ産業」は高い重要度が割り当てられる。また、ルール703より、「締め切られた」は目的語を伴わない動詞なので、重要度は低くなる。

【0028】同様に第2文では、ルール702により、具体数値である「十倍強に当たる五百万一五百万件になった」の重要度が高く判定される。同様に第3文では、ルール701により、企業名「日本テレコム」の重要度が高くなるが、同時に固有名詞は新しく出現するたびに重要度が低くなることから、「日本たばこ産業株」よりも低い重要度が割り当てられる。動詞「下回った」は目的語「事前の予想を」を伴っていることから、ルール703により、重要度が高くなる。

【0029】このようにして、1要素ごとに重要度を付与した結果を、出力手段604によって出力する。

【0030】次に1要素ごとに重要度を付与した結果に対し、図1の重要要素抽出部108を用いて重要要素を抽出する。重要要素抽出部108を図8を用いて説明する。1要素ごとに重要度を付与した数文に対し、1文中の主語抽出801、動詞抽出802、その他の修飾語抽出803を分けて行い、それを各文について行なう。

【0031】次に、手段804にて、抽出された各要素に対し、要素と重要度とを取り出してテーブル化を行ない、手段805にて各テーブルごとに重要度に従ってソートを行なう。手段806にて、この結果と、予め入力されてある各要素に対する要約パラメタ807とを用いて、上位から任意数の要素を抽出する。このとき、主語、動詞は必ずゼロにならないようにパラメタを設定する必要がある。これによって、各文ごとに抽出された重要要素、主語・目的語・動詞、もしくは主語・動詞が得られる。

【0032】例文の場合で説明する。まず、1文目の主語は「購入申し込みが」、動詞は「締め切られた」、修飾語は「日本たばこ産業株の」「一般売り出しの」「八日で」が抽出され、それぞれに対して重要度と要素が対になったテーブルが作成されている。この場合、主語、動詞は1要素ずつしかないが、修飾語は重要度の高さから「日本たばこ産業の」「一般売り出しの」「八日で」の順にソートされる。ここで、予め入力されてある要素に対する要約パラメタ807を、各テーブルごとに上位1要素のみを抽出するように設定してあるとすると、手段806によって、主語は「購入申し込みが」、動詞は「締め切られた」、修飾語は「日本たばこ産業株の」が抽出される。

【0033】同様に2文目は、主語は「申し込み件数は」、動詞は「なった」、修飾語は「十倍強に当たる五百万一五百万件に」が抽出される。

【0034】同様に3文目は、主語は「申し込み件数は」、動詞は「下回った」、修飾語は「事前の予想を」が抽出される。

【0035】こうして抽出された各要素は、図1の要約文生成部109によって要約文に生成される。これを図9により示す。例文の場合、各文ごとに抽出された重要要素は、手段901によって文法に従って並び替えを行ない、次の3つの要約文が生成される。

【0036】「日本たばこ産業株の購入申し込みが、締め切られた」「申し込み件数は、十倍強に当たる五百万一五百万件になった」、「申し込み件数は、事前の予想を下回った」。

【0037】次に、要素の重なりを縮約する手段902により、2文目主語「申し込み件数は」と、3文目主語「申し込み件数は」を縮約し、「申し込み件数は、十倍強に当たる五百万一五百万件になり、事前の予想を下回った」という文を生成する。以上の流れにより、要約文生成部109（図1）は、「日本たばこ産業株の購入申し込みが、締め切られた」「申し込み件数は、十倍強に当たる五百一五百万件になり、事前の予想を下回った」の2文を要約文として出力する。

【0038】以上、図1において、出力装置110によって上記要約文を出力し要約文111を得る。以上述べた装置が、請求項（1）記載の要約装置である。

【0039】以下、請求項(2)記載の要約装置について説明する。

【0040】図1において、要約文生成部109の出力結果と、並べ替えルールテーブル113を用いて文の並べ替えを行なう並べ替え部112を、図10および図11を用いて説明する。要約文生成部109によって出力された各要約文に対し、手段1201にて、各要素の重要度の総和が最も高い1文を、要約文の先頭に並べ替える。次に、各要素に付与されている解析情報から意味つながりを調べ、つながりのある文ごとに並べ替える。例文の場合、もっとも重要度の高い要約文は第1文なので、並び替えを行なわない。

【0041】並べ替え部112の結果とそれを出力する装置114によって自然言語要約文115を得る装置が請求項(2)記載の要約装置である。例文の場合、「日本たばこ産業株の購入申し込みが、締め切られた」「申し込み件数は、十倍強に当たる五百一十五万五千件になり、事前の予想を下回った」の2文が要約文として得られる。

【0042】以下、請求項(3)記載の要約装置について説明する。

【0043】図1において、上記並べ替え部112の結果を機械翻訳機116によって自動翻訳し、出力装置117を用いて第2自然言語要約文を得る装置が、請求項(3)記載の要約装置である。例文の場合に日英翻訳を適用すると、“The purchase application of Japan cigarette industry stocks was closed.” “The application number of cases reached 5 - 5.5 million cases that amount to little over ten times and dropped below a beforehand expectation.” の2文が得られる。

【0044】上記において、記事要約装置および記事要約処理方法について説明したが、当該記事要約処理方法はプログラムの形で記述して記録媒体に保持することができる。したがって、本発明の権利範囲は当該記録媒体を含むものとする。

【0045】

【発明の効果】以上説明した如く、本発明によれば、要約ルールを用いることにより、従来の重み付けや統計処

* 理のコストを省くことができる。また、文単位だけではなく要素単位の要約を行なえる。また、他言語に要約することが可能な翻訳装置である。

【図面の簡単な説明】

【図1】本要約装置の基本ブロックである。

【図2】実施例に用いる入力例文である。

【図3】本発明の重要度判定部の処理フローである。

【図4】重要度判定部で使用される重要度判定ルールテーブルである。

【図5】本発明の重要文抽出部の処理フローである。

【図6】本発明の重要要素判定部の処理フローである。

【図7】重要要素判定部で使用される重要要素判定ルールテーブルである。

【図8】本発明の重要要素抽出部の処理フローである。

【図9】本発明の要約文生成部の処理フローである。

【図10】本発明の並べ替え部の処理フローである。

【図11】並べ替え部で使用される並べ替えルールテーブルである。

【符号の説明】

100：第1自然言語記事

101：要約パラメタ

102：入力装置

103：重要度判定部

104：重要度判定ルールテーブル

105：重要文抽出部

106：重要要素判定部

107：重要要素判定ルールテーブル

108：重要要素抽出部

109：要約文生成部

30 110：出力装置

111：第1自然言語要約文

112：並べ替え部

113：並べ替えルールテーブル

114：出力装置

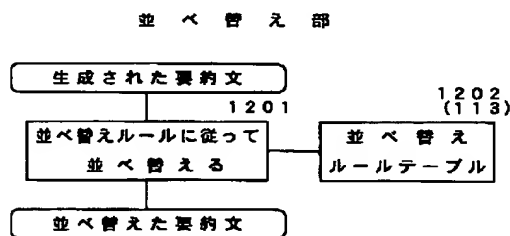
115：第1自然言語要約文

116：機械翻訳機

117：出力装置

118：第2自然言語要約文

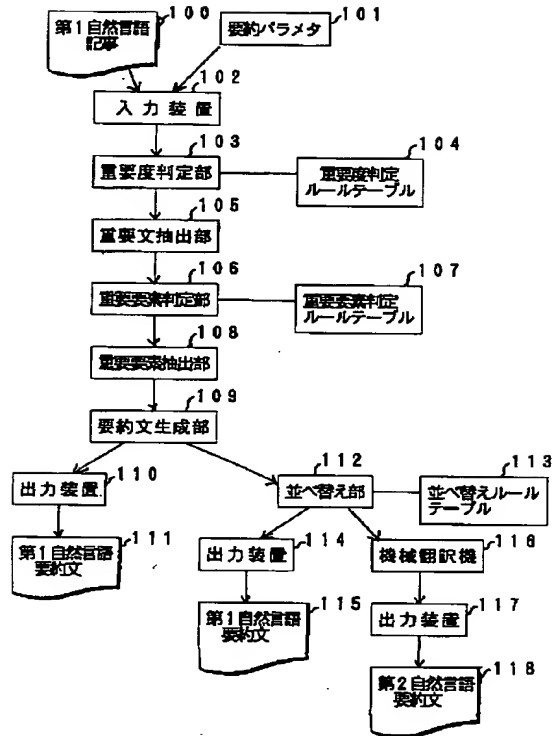
【図10】



【図11】

並べ替えルールテーブル	
重要度	条 件
1	最も重要度の高い文を第1文にする。
2	関連のある話題をまとめる。
...

【図1】

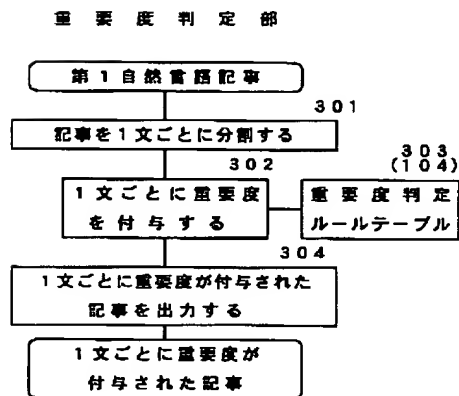


【図2】

入力文

- 1文目 日本たばこ産業（ＪＴ）株の一般売り出しの購入申し込みが、八日で締め切られた。
- 2文目 市場関係者によると、申込件数は売出株数の十倍強に当たる五百万～五百五十万件になった模様だ。
- 3文目 六日上場した日本テレコムの子株が公募価格を割り込んでいることが影響、申込件数は事前の予想を下回った。
- 4文目 ＪＴ株の売出株数は四十三万六千六百六十六株。
- 5文目 購入申し込みの件数は、一千万件を上回るとの観測もあったが、日本電信電話（ＮＴＴ）株（第一次放出百六十五万株、売出価格百十九万七千円）の千五十九万件、東日本旅客鉄道（ＪＲ東日本）株（百四十万株、同三十八万円）の千四十八万件に比べ半分程度の水準となった。

【図3】

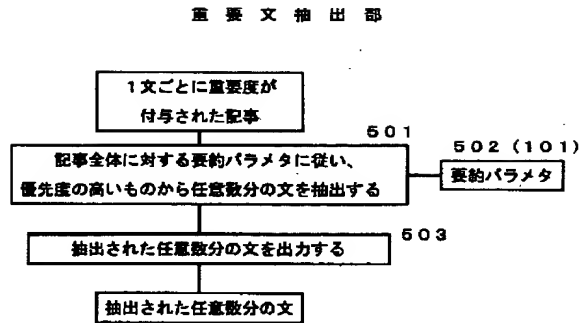


【図4】

重要度判定ルールテーブル

重要度	条件	
...	
D001	第1段落第1文目の文である。	401
D002	第1段落第2文目の文である。	
...	
A001	企業・官庁名が含まれる。	402
A002	個人名が含まれる。	
A003	固有名詞が含まれる。	403
A004	数値が含まれる。	
A005	固有名詞が複数出現する場合、重要度の高さは出現順。	404
...	
9001	「～そうだ」「～ようだ」は重要度が低い。	405
...	

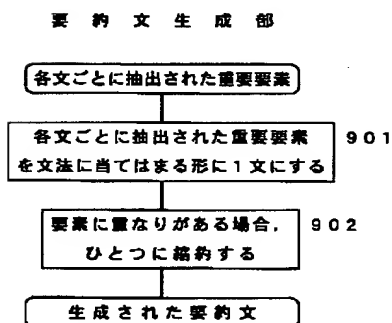
【図5】



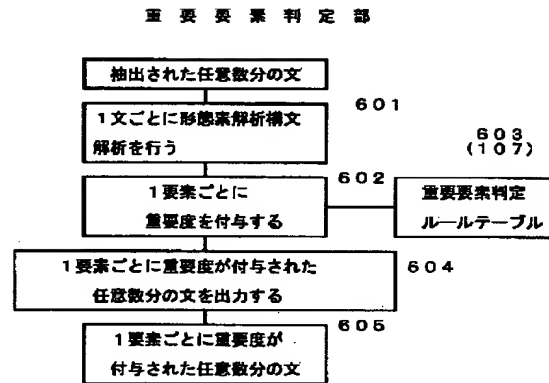
【図7】

重要要素判定ルールテーブル		
重要度	対象	条件
...
E001	名詞句	企業名、官庁名、個人名を含む。
E002	名詞句	固有名称が複数出現するとき、出現の早いものほど重要度が高くなる。
E003	名詞句	具体的数値を含む。
E004	名詞句	具体名詞である。
...
A001	動詞	目的語を伴う動詞の方が重要度が高い。
9001	動詞	「動詞+ようだ」ではない。
...
X001	*	修飾語の重要度は低い。
X002	*	() の内部は省略する。
...

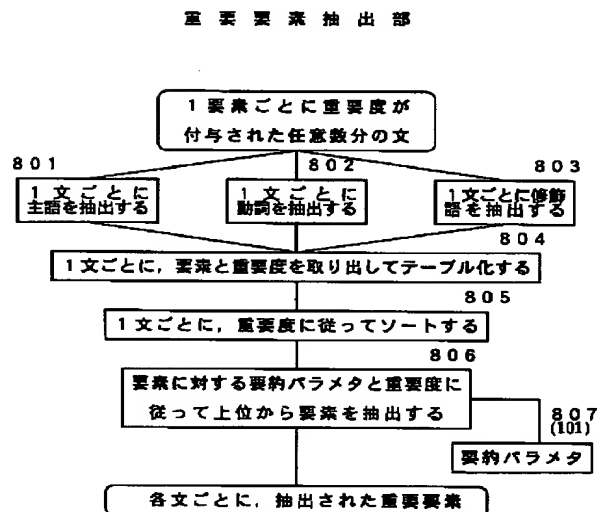
【図9】



【図6】



【図8】



**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☐ **FADED TEXT OR DRAWING**
- ☒ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.